

Exposure to Dust - Logistic Regression and Search for Outliers

November 1, 2024

First of all, the dust data are loaded:

```
> library(catdata)
> data(dust)
> attach(dust)
```

First, the subsample of non-smokers is considered. A main effect logit model yields the following results:

```
> dustlogitnon1=glm(bronch ~ dust+years, family=binomial, data=dust[(dust$smoke==0),])
> summary(dustlogitnon1)
```

The same model as above is used without observation 1245 which can be regarded as an outlier:

```
> dustlogitnon2 <- glm(bronch ~ dust+years, family=binomial,
+                       data=dust[(dust$smoke==0)&(dust$dust<10),])
> summary(dustlogitnon2)
```

The following calculations are based on the complete dataset. Therefore, main effect logit models are fitted for all observations and without observation 1246, respectively:

```
> dustlogit1 <- glm(bronch ~ dust+years+smoke, family=binomial, data=dust)
> summary(dustlogit1)

> dustlogit2 <- glm(bronch ~ dust+years+smoke, family=binomial,
+                   data=dust[(dust$dust<20),])
> summary(dustlogit2)
```